Designation: E2617 – 10

Standard Practice for Validation of Empirically Derived Multivariate Calibrations[1]

This standard is issued under the fixed designation E2617; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice covers requirements for the validation of empirically derived calibrations (Note 1) such as calibrations derived by Multiple Linear Regression (MLR), Principal Component Regression (PCR), Partial Least Squares (PLS), Artificial Neural Networks (ANN), or any other empirical calibration technique whereby a relationship is postulated between a set of variables measured for a given sample under test and one or more physical, chemical, quality, or membership properties applicable to that sample.

NOTE 1—Empirically derived calibrations are sometimes referred to as "models" or "calibrations." In the following text, for conciseness, the term "calibration" may be used instead of the full name of the procedure.

1.2 This practice does not cover procedures for establishing said postulated relationship.

1.3 This practice serves as an overview of techniques used to verify the applicability of an empirically derived multivariate calibration to the measurement of a sample under test and to verify equivalence between the properties calculated from the empirically derived multivariate calibration and the results of an accepted reference method of measurement to within control limits established for the prespecified statistical confidence level.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E131 Terminology Relating to Molecular Spectroscopy

E1655 Practices for Infrared Multivariate Quantitative Analysis

E1790 Practice for Near Infrared Qualitative Analysis

3. Terminology

3.1 For terminology related to molecular spectroscopic methods, refer to Terminology E131. For terminology related to multivariate quantitative modeling refer to Practices E1655. While Practices E1655 is written in the context of multivariate spectroscopic methods, the terminology is also applicable to other multivariate technologies.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *accuracy*—the closeness of agreement between a test result and an accepted reference value.

3.2.2 *bias*—the arithmetic average difference between the reference values and the values produced by the analytical method under test, for a set of samples.

3.2.3 *detection limit*—the lowest level of a property in a sample that can be detected, but not necessarily quantified, by the measurement system.

3.2.4 *estimate*—the constituent concentration, identification, or other property of a sample as determined by the analytical method being validated.

3.2.5 *initial validation*—validation that is performed when an analyzer system is initially installed or after major maintenance.

3.2.6 *Negative Fraction Identified*—the fraction of samples not having a particular characteristic that is identified as not having that characteristic.

3.2.6.1 *Discussion*—Negative Fraction Identified assumes that the characteristic that the test measures either is or is not present. It is not applicable to tests with multiple possible outcomes.

3.2.7 *ongoing periodic revalidation*—the quality assurance process by which, in the case of quantitative calibrations, the bias and precision or, in the case of qualitative calibrations, the Positive Fraction Identified and Negative Fraction Identified performance determined during initial validation are shown to be sustained.

3.2.8 *Positive Fraction Identified*—the fraction of samples having a particular characteristic that is identified as having that characteristic.

3.2.8.1 *Discussion*—Positive Fraction Identified assumes

[1] This practice is under the jurisdiction of ASTM Committee E13 on Molecular Spectroscopy and Separation Science and is the direct responsibility of Subcommittee E13.11 on Multivariate Analysis.

Current edition approved March 1, 2010. Published April 2010. Originally approved in 2008. Last previous edition approved in 2009 as E2617 – 09a. DOI: 10.1520/E2617-10.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

that the characteristic that the test measures either is or is not present. It is not applicable to tests with multiple possible outcomes.

3.2.9 *precision*—the closeness of agreement between independent test results obtained under stipulated conditions.

3.2.9.1 *Discussion*—Precision may be a measure of either the degree of reproducibility or degree of repeatability of the analytical method under normal operating conditions. In this context, reproducibility refers to the use of the analytical procedure in different laboratories, as in a collaborative study.

3.2.10 *quantification limit*—the lowest level of a sample property which can be determined with acceptable precision and accuracy under the stated experimental conditions.

3.2.11 *range*—the interval between the upper and lower levels of a property (including these levels) that has been demonstrated to be determined with a suitable level of precision and accuracy using the method as specified.

3.2.12 *reference value*—the metric of a property as determined by well-characterized method, the accuracy of which has been stated or defined, that is, another, already-validated method.

3.2.13 *validation*—the statistically quantified judgment that an empirically derived multivariate calibration is applicable to the measurement on which the calibration is to be applied and can perform property estimates with, in the case of quantitative calibrations, acceptable precision, accuracy and bias or, in the case of qualitative calibrations, acceptable Positive Fraction Identified and Negative Fraction Identified, as compared with results from an accepted reference method.

3.2.14 *validation space*—the region(s) of a calibration's multivariate sample space populated by the independent validation samples which are used to validate the calibration.

4. Summary of Practice

4.1 Validating an empirically derived multivariate calibration (model) consists of four major procedures: validation at initial development, revalidation at initial deployment or after a revision, ongoing periodic revalidation, and qualification of each measurement before using the calibration to estimate the property(s) of the sample being measured.

5. Significance and Use

5.1 This practice outlines a universally applicable procedure to validate the performance of a quantitative or qualitative, empirically derived, multivariate calibration relative to an accepted reference method.

5.2 This practice provides procedures for evaluating the capability of a calibration to provide reliable estimations relative to an accepted reference method.

5.3 This practice provides purchasers of a measurement system that incorporates an empirically derived multivariate calibration with options for specifying validation requirements to ensure that the system is capable of providing estimations with an appropriate degree of agreement with an accepted reference method.

5.4 This practice provides the user of a measurement system that incorporates an empirically derived multivariate calibration with procedures capable of providing information that may be useful for ongoing quality assurance of the performance of the measurement system.

5.5 Validation information obtained in the application of this practice is applicable only to the material type and property range of the materials used to perform the validation and only for the individual measurement system on which the practice is completely applied. It is the user's responsibility to select the property levels and the compositional characteristics of the validation samples such that they are suitable to the application. This practice allows the user to write a comprehensive validation statement for the analyzer system including specific limits for the validated range of application and specific restrictions to the permitted uses of the measurement system. Users are cautioned against extrapolation of validation results beyond the material type(s) and property range(s) used to obtain these results.

5.6 Users are cautioned that a validated empirically derived multivariate calibration is applicable only to samples that fall within the subset population represented in the validation set. The estimation from an empirically derived multivariate calibration can only be validated when the applicability of the calibration is explicitly established for the particular measurement for which the estimation is produced. Applicability cannot be assumed.

6. Methods and Considerations

6.1 When validating an empirically derived multivariate calibration, it is the responsibility of the user to describe the measurement system and the required level of agreement between the estimations produced by the calibration and the accepted reference method(s).

6.2 When validating a measurement system incorporating an empirically derived multivariate calibration, it is the responsibility of the user to satisfy the requirements of any applicable tests specific to the measurement system including any Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ) requirements; which may be mandated by competent regulatory authorities, an applicable Quality Assurance (QA), or Standard Operating Procedure (SOP) or be recommended by the instrument or equipment manufacturer.

6.3 *Reference Values and Quality Controls for the Accepted Reference Method:*

6.3.1 The reference (or true) value which is compared with each respective estimate produced by the empirically derived multivariate calibration is established by applying an accepted reference method, the characteristics of which are known and stated, to the sample from which the measurement system derives the measurement.

6.3.2 To ensure the reliability of the reference values provided by an accepted reference method, appropriate quality controls should be applied to the accepted reference method.

7. Procedure

7.1 The objective of the validation procedure is to quantify the performance of an empirically derived multivariate calibration in terms of, in the case of quantitative calibrations, precision, accuracy and bias or, in the case of qualitative calibrations, Positive Fraction Identified and Negative Fraction Identified relative to an accepted reference method for each property of interest. The user must specify, based on the intended use of the calibration, acceptable precision and bias or Positive Fraction Identified and Negative Fraction Identified performance criteria before initiating the validation. These criteria will be dependent on the intended use of the analyzer and may be based, all or in part, on risk based criteria.

7.1.1 The acceptable performance criteria specified by the user may be constant over the entire range of sample variability. Alternatively, different acceptable performance criteria may be specified by the user for different sub-ranges of the full sample variability.

7.2 Validation of calibration is accomplished by using the calibration to estimate the property(s) of a set of validation samples and statistically comparing the estimates for these samples to known reference values. Validation requires thorough testing of the model with a sufficient number of representative validation samples to ensure that it performs adequately over the entire range of possible sample variability.

7.3 *Initial Validation Sample Set:*

7.3.1 For the initial validation of a multivariate model, an ideal validation sample set will:

7.3.1.1 Contain samples that provide sufficient examples of all combinations of variation in the sample properties which are expected to be present in the samples which are to be analyzed using the calibration;

7.3.1.2 Contain samples for which the ranges of variation in the sample properties is comparable to the ranges of variation expected for samples that are to be analyzed using the model;

7.3.1.3 Contain samples for which the respective variations of the sample properties are uniformly and mutually independently distributed over their full respective ranges or, when applicable, subranges of variation; and

7.3.1.4 Contain a sufficient number of samples to statistically test the relationships between the measured variables and the properties that are modeled by the calibration.

7.3.2 For simple systems, sufficient validation samples can generally be obtained to meet the criteria in 7.3.1.1-7.3.1.4. For complex mixtures, obtaining an ideal validation set may be difficult if not impossible. In such cases, it may be necessary to validate discrete subranges of the calibration incrementally, over time as samples become available.

7.3.3 The number of samples needed to validate a calibration depends on the complexity of the calibration, the ranges of property variation over which the calibration is to be applied, and the degree of confidence required. It is important to validate a calibration with as many samples as possible to maximize the likelihood of challenging the calibration with rarely occurring, but potentially troublesome samples. The number and range of validation samples should be sufficient to validate the calibration to the statistical degree of confidence required for the application. In all cases, a minimum of 20 validation samples is recommended. In addition, the validation samples should:

7.3.3.1 Multivariately span the ranges of sample property values over which the calibration will be used; that is, the span and the standard deviation of the ranges of sample property values for the validation samples should be at least 100 % of the spans of the sample property values over which the calibration will be used, and the sample property values for the validation samples should be distributed as uniformly as possible throughout their respective ranges, and the variations of the sample property values among the samples should be as mutually independent as possible; and

7.3.3.2 Span the ranges of the independent variables over which the calibration will be used; that is, if the range of an independent variable is expected to vary from a to b, and the standard deviation of the independent variable is c, then the variations of that independent variable in the set of validation samples should cover at least 100 % of the range from a to b, and should be distributed as uniformly as possible across the range such that the standard deviation in that independent variable estimated for the validation samples will be at least 95 % of c.

(1) When validating a calibration for which detection limit or quantification limit is an important consideration, the user should include a number of validation samples whose property(s) are close to the detection or quantification limit(s) sufficient to validate the respective limit(s) to the statistical degree of confidence required for the application.

7.4 For quantitative calibrations, the validation error for each property in each sample is given by the Standard Error of Validation (SEV) and bias for that property.

7.4.1 The validation bias, ev-, is a measure of the average difference between the estimates made based on the empirical model and the results obtained on the same validation samples using the reference method.

7.4.1.1 If there are single reference values and estimates for each validation sample, the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v} (\hat{v}_i - v_i)}{v} \quad (1)$$

where:

$\hat{v}_i$ = estimate from the model for the ith sample,
v_i = accepted reference value for the ith sample, and
v = number of validation samples.

7.4.1.2 If replicate estimates and a single reference value are available for the validation samples, then the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v} \sum_{j=1}^{r_i} (\hat{v}_{ij} - v_i)}{\sum_{i=1}^{v} r_i} \quad (2)$$

where:

$\hat{v}_{ij}$ = the jth estimate for the ith validation sample, and

r_i = number of replicate estimates for the ith validation sample.

7.4.1.3 If a single estimate and multiple reference values are available for the validation samples, then the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_i - v_{ij})}{\sum_{i=1}^{v} s_i} \quad (3)$$

where:

$\hat{v}_i$ = estimate for the ith validation sample,
v_{ij} = the jth reference value for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

7.4.1.4 If multiple estimates and multiple reference values are available for the validation samples, then the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\sum_{k=1}^{s_i}(\hat{v}_{ij} - v_{ik})}{\sum_{i=1}^{v} r_i s_i} \quad (4)$$

where:

$\hat{v}_{ij}$ = the jth estimate for the ith validation sample,
v_{ik} = the kth reference value for the ith validation sample,
r_i = number of replicate estimates for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

7.4.2 The SEV, also called the Standard Error of Prediction (SEP) and the Standard Deviation of Validation Residuals (SDV), are measures of the expected agreement of the empirical model and the reference method. The calculation of SEV and SDV depend on whether replicate estimates or reference values, or both, are used.

7.4.2.1 If there are single reference values and estimates for each validation sample, then SEV and SDV are calculated as:

$$\text{SEV} = \sqrt{\frac{\sum_{i=1}^{v}(\hat{v}_i - v_i)^2}{v}} \quad (5)$$

$$\text{SDV} = \sqrt{\frac{\sum_{i=1}^{v}(\hat{v}_i - v_i - \bar{e}_v)^2}{v}}$$

where:

$\hat{v}_i$ = estimate from the model for the ith sample,
v_i = accepted reference value for the ith sample, and
v = number of validation samples.

7.4.2.2 If replicate estimates and a single reference value are available for the validation samples, then SEV and SDV are calculated as:

$$\text{SEV} = \sqrt{\frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_{ij} - v_j)^2}{\sum_{i=1}^{v} r_i}} \quad (6)$$

$$\text{SDV} = \sqrt{\frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_{ij} - v_i - \bar{e}_v)^2}{\sum_{i=1}^{v} r_i}}$$

where:

$\hat{v}_{ij}$ = the jth estimate for the ith validation sample, and
r_i = number of replicate estimates for the ith validation sample.

NOTE 2—If each validation sample is estimated r times, an average estimate could be used in 7.4.2.1, but then the SEV calculated would represent the expected agreement between the average of r estimations and a single reference measurement, not the agreement based on a single estimation from the empirical model.

7.4.2.3 If a single estimate and multiple reference values are available for the validation samples, then SEV and SDV are calculated as:

$$\text{SEV} = \sqrt{\frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_i - v_{ij})^2}{\sum_{i=1}^{v} s_i}} \quad (7)$$

$$\text{SDV} = \sqrt{\frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_i - v_{ij} - \bar{e}_v)^2}{\sum_{i=1}^{v} s_i}}$$

where:

$\hat{v}_i$ = estimate for the ith validation sample,
$\hat{v}_{ij}$ = the jth reference value for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

NOTE 3—If each validation sample has s reference values, an average estimate could be used in 7.4.2.1, but then the SEV calculated would represent the expected agreement between an estimate from the empirical model and the average of s reference measurements, not a the agreement relative to a single reference measurement.

7.4.2.4 If multiple estimates and multiple reference values are available for the validation samples, then SEV and SDV are calculated as:

$$\text{SEV} = \sqrt{\frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\sum_{k=1}^{s_i}(\hat{v}_{ij} - v_{ik})^2}{\sum_{i=1}^{v} r_i s_i}} \quad (8)$$

$$\text{SDV} = \sqrt{\frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\sum_{k=1}^{s_i}(\hat{v}_{ij} - v_{ik} - \bar{e}_v)^2}{\sum_{i=1}^{v} r_i s_i}}$$

where:

$\hat{v}_{ij}$ = the jth estimate for the ith validation sample,
$\hat{v}_{ik}$ = the kth reference value for the ith validation sample,
r_i = number of replicate estimates for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

NOTE 4—If each validation sample has r estimates and s reference values, average estimates and reference values could be used in 7.4.1.1, but then the SEV calculated would represent the expected agreement between r estimates from the empirical model and the average of s reference measurements, not a the agreement between a single estimate and reference measurement.

7.4.3 *Significance of Validation Bias*—A t-value can be calculated as:

$$t = \frac{e_v d_v}{\text{SDV}} \quad (9)$$

where:

d_v = degrees of freedom and is equal to the denominator in the bias calculation.

NOTE 5—The t-value is compared to a critical t-value for the desired probability level (typically 95 %).

7.4.3.1 If the calculated t-value is less than the critical t-value, then the validation bias is not statistically significant and the empirical model and reference method are expected to on average yield the same result. In this case, either SEV or SDV are adequate measures of the expected agreement between the empirical model and the reference method. If the validation bias is of practical significance relative to the user specified bias requirement, then the precision of the empirical model results is insufficient to achieve the user requirement.

7.4.3.2 If the calculated t-value is greater than the critical t-value, then the validation bias is statistically significant. In this case SDV is a better measure of the expected agreement between the results of the empirical model and the reference method. While the bias may be statistically significant, it may not be of practical significance relative to the user specified requirements for the empirical model.

7.5 *Positive and Negative Fractions Identified:*

7.5.1 The Positive Fraction Identified of the calibration is given by: Positive Fraction Identified = (number of samples identified as having a stated characteristic) / (total number of samples having the stated characteristic).

7.5.2 The Negative Fraction Identified of the calibration is given by: Negative Fraction Identified = (number of samples identified as not having a stated characteristic) / (total number of samples not having the stated characteristic).

7.5.3 The equations for Positive Fraction Identified and Negative Fraction Identified assume that the characteristic being measured either is or isn't present. It is not applicable to tests with multiple possible outcomes.

7.6 The users should use statistical tests and decision criteria appropriate to the application to decide if the SEV and bias are within statistically acceptable limits.

7.7 *Samples for Revalidation After Initial Deployment and Ongoing Periodic Revalidation Samples:*

7.7.1 The user must determine, based on the particulars of each application, the appropriate timing and number of samples required for revalidation after initial deployment and for ongoing periodic revalidation.

7.7.1.1 The timing and number of revalidation samples may be adjusted from time to time as experience is gained in applying the calibration under actual conditions.

7.7.1.2 In many cases revalidation samples are restricted to "samples of opportunity" and limited to samples from actual production operations. In such cases, care should be taken to schedule revalidation samples as asynchronously as possible with respect to recurring conditions such as time of day, production process operating conditions, phase or stage of production process, ambient conditions, operating personnel, etc. This listing of potential conditions for consideration is exemplary, not comprehensive; the user should take into account any external conditions pertinent to the application.

7.7.2 It is recommended that the results of ongoing periodic revalidation should be monitored or tracked by control charting.

8. Qualification of Each Measurement Prior to Application of the Validated Calibration

8.1 The independent variables measured from a sample under test must be evaluated to ensure that this measurement is eligible to be processed by the calibration to produce estimates of the property(s) of interest. The purpose of this eligibility test is to determine, within user specified statistical limits, if the validation samples used to validate the calibration are sufficiently representative of (similar to) the sample under test. In other words, the purpose of this step is to confirm that the measurement from the sample under test is within the calibration's validation space. If the measurement is eligible, the estimates should fall within accuracy and precision bounds determined during the validation. If the measurement is not eligible, then the accuracy and precision of the estimates are not known based on the validation. The measurement of a sample under test may be tested for eligibility using Mahalanobis distance, Nearest Neighbor Mahalanobis Distance (NNMD), or Standard Residual Variance in the Independent Variables (SRVIV), either singly or in combination. The user may also specify additional eligibility criteria if and as appropriate to the application.

8.1.1 The development of an empirical model will typically involve transformation of the independent variables. By way of illustration, such transformation may include one or more of the following:

8.1.1.1 Linearization of the independent variables (for example, conversion from transmission to absorbance, from reflectance to log(1-reflectance), etc.);

8.1.1.2 Digital filtering (smoothing, digital derivatives);

8.1.1.3 Orthogonalization (Orthogonal Signal Correction);

8.1.1.4 Rank reduction (Principal Components Analysis (PCA) or PLS);

8.1.1.5 Squares, cross products or nonlinear functions of variables;

8.1.1.6 Explicit artifact removal (cosmic ray event removal);

8.1.1.7 Centering or baseline correction;

8.1.1.8 Arbitrary scaling, variance scaling, or auto scaling;

8.1.1.9 Exclusion of one or more independent variables from use in the calibration; and

8.1.1.10 Integration of peaks with or without baseline correction.

8.1.2 Mahalanobis distance, NNMD, and SRVIV statistics are calculated after applying the same transformations to the measurement being qualified which were applied to the measurements used to produce and validate the calibration.

8.2 SRVIVs can sometimes be employed to determine if the samples used to validate the empirical model are sufficiently representative of (similar to) the sample under test. SRVIV is intended to detect any anomalous variance which may be present in the measurement from new signals (for example, new chemical components, new instrumental or sample conditions, etc.) that were not represented in the validation samples. If the validation samples are sufficiently representative of the (unknown) sample under test, then the amount of residual variance in the independent variables of the sample under test will be statistically indistinguishable from the amount of residual variance in the validation samples. This is always a necessary criterion for qualification testing, but it may not always be solely sufficient. If the empirical calibration utilizes most of the non-noise portion of variance in the independent variables, the residual variance will be a very sensitive measure of any aberrant variance present in the data for the sample under test. Alternatively, if the empirical model is based on a small fraction of the non-noise portion of the variance in the independent variable, then tests based on the statistics of the SRVIV are unlikely, used alone, to provide adequate warning of measurements, which are not qualified for estimation by the calibration.

8.2.1 The residual variance in the independent variables is defined as that fraction of the variance in the variables which is not spanned by the validation samples' basis space comprising an appropriate number of abstract factors determined by either PCA (**1, 2**)[3] or PLS (**1, 2**). If the $f \times v$ matrix $\mathbf{X}$ comprises column vectors, each of which contains the f independent variables (for example, the spectrum) of v validation samples; the $f \times k$ matrix $\mathbf{P}$ comprises column vectors, each of which contains one of the factors comprising the PCA or PLS basis space; the $f \times k$ matrix $\mathbf{T}$ comprises the scores of the v validation samples for the k basis vectors; the $f \times v$ matrix $\hat{\mathbf{X}}$ comprises column vectors, each of which contains the reconstructions of respective columns of $\mathbf{X}$ by the k factors comprising the basis space in $\mathbf{P}$; and the $f \times v$ matrix $\mathbf{R}$ comprises column vectors, each of which contains the residual variance in each corresponding measurement in $\mathbf{X}$ which is not spanned by the basis space; then:

$$\begin{aligned}\mathbf{R} &= \mathbf{X} - \hat{\mathbf{X}} \\ &= \mathbf{X} - \mathbf{P}\mathbf{T}^T \\ &= \mathbf{X} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{X}\end{aligned} \tag{10}$$

and the standard residual (SRVIV) is then given by:

$$\text{SRVIV} = \sqrt{\frac{\sum_{i=1}^{f}\sum_{j=1}^{v} R_{ij}^2}{f \cdot k}} \tag{11}$$

where:

T = matrix transpose, and

$(f \cdot k)$ = number of degrees of freedom of the residuals.

For PCA, an alternative method of calculating the residual variance uses the loadings $\mathbf{L}$, singular values, $\sum$, and scores, $\mathbf{S}$, from the singular value decomposition (SVD) (see Practices E1655) of $\mathbf{X}$. The equation for SVD is:

$$\mathbf{X} = \mathbf{L}_v \sum_v \mathbf{S}_v^T \tag{12}$$

If the $f \times k$ matrix $\mathbf{L}_v$ comprises column vectors, each of which contains one of the SVD factors (loadings) comprising the basis space; the $k \times k$ matrix $\sum_v$ is a diagonal matrix which contains the k singular values for the respective factors in $\mathbf{L}_v$; the $v \times k$ matrix $\mathbf{S}$ comprises column vectors, each of which contains the k scores of each validation sample in $\mathbf{X}$ against the respective factors in $\mathbf{L}_v$; and the $f \times v$ matrix $\mathbf{R}$ comprises column vectors, each of which contains the residual variance in each corresponding measurement in $\mathbf{X}$ which is not spanned by the basis space; then:

$$\mathbf{R} = \mathbf{X} - \hat{\mathbf{X}} \tag{13}$$

where:

$\hat{\mathbf{X}} = \mathbf{L}_v \sum_v \mathbf{S}_v^T$

The standard residual is then given by Eq 11.

NOTE 6—Relative to the notation for PCA, typically either $\mathbf{L}_v \sum_v = \mathbf{P}$ and $\mathbf{T} = \mathbf{S}_v$ or $\mathbf{L}_v = \mathbf{P}$ and $\mathbf{S}_v \sum_v = \mathbf{T}$.

8.2.1.1 For the purposes of this calculation, the user should determine that all rows of the matrix of residuals have equal variance and that all columns of the matrix of residuals have equal variance before applying the equation for standard residual.

8.2.1.2 If the abstract factors for the basis space were calculated using PLS, Eq 15 is an approximation in that it does not account for the fact that, in general, each additional PLS factor does not reduce the degrees of freedom of the system by exact integer amount.

8.2.1.3 PLS algorithms, depending upon how they are implemented in software, can produce either orthogonal PLS factors (often called the PLS loading weights) or non-orthogonal PLS factors (often called the PLS loadings). Either type of factors may be used for the basis space, but each type of factor may yield different qualification results. The orthogonal PLS loading weights will usually produce a standard residual for the validation samples which is very close in magnitude to the standard residual produced by the corresponding number of PCA factors. The non orthogonal PLS loadings will usually produce a standard residual for the validation samples which is slightly or significantly larger than the standard residual produced by the corresponding number of PCA factors. Accordingly, the SRVIV test when using the PLS

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.

loading weights may provide more sensitive detection of unqualified samples in some cases.

8.2.2 After the SRVIV is calculated for the validation samples, a user specified confidence limit is then applied to the SRVIV to establish a range of expected SRVIV values for samples that are well represented by the validation samples. Based on the user specified confidence limit users establish an upper cutoff for the SRVIV for samples under test.

8.2.3 If the column vector, $\mathbf{x}_u$, contains the measurement of the sample under test, the residual, $\mathbf{r}$, is given by:

$$\begin{aligned}\mathbf{r} &= \mathbf{x}_u - \hat{\mathbf{x}}_u \\ &= \mathbf{x}_u - \mathbf{P}\mathbf{t} \\ &= \mathbf{x}_u - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{x}_u\end{aligned} \tag{14}$$

and the standard residual, SRVIV, is then given by:

$$\text{SRVIV} = \sqrt{\frac{\sum_{i=1}^{f} r_i^2}{f}} \tag{15}$$

where:

f = number of rows in $\mathbf{x}_u$.

8.2.4 If the SRVIV value for the sample under test is less than the cutoff value established by the user in accordance with 8.2.2, then the measurement of the sample under test is not disqualified, with respect to residual variance in the independent variables, for estimation by the calibration. In such cases, the validation samples are sufficiently representative of the sample under test that estimates should fall within the accuracy and precision bounds determined during the validation.

8.2.5 If the SRVIV value for the sample under test is greater than the user established cutoff value, then the measurement of the sample under test is not qualified, with respect to residual variance in the independent variables, for estimation by the calibration. Validation samples are not sufficiently representative of the sample under test and the accuracy and precision of the estimates are not known based on the validation.

8.3 The Mahalanobis distance can indicate whether or not the transformed variables for a sample under test fall within the multivariate space defined by the transformed variables for the validation sample set such that the sample under tests is an interpolated rather than an extrapolated sample with respect to the samples in the validation set. If the validation samples are sufficiently representative of the sample under test, then the Mahalanobis distance of the measurement with respect to the centroid of the validation samples will be statistically indistinguishable from the Mahalanobis distances of the set of validation samples.

8.3.1 The Mahalanobis distance, h, of the sample under test is given by:

$$h = \mathbf{x}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{x} \tag{16}$$

where:

$\mathbf{x}$ = the mean centered vector containing the measured values for the sample under test, and

$\mathbf{X}$ = the mean centered matrix containing the measured values of the validation samples.

NOTE 7—Mean centering is accomplished by calculating the mean (average) spectrum of all of the calibration samples and subtracting this mean spectrum from each validation sample and from the sample under test prior to calculating h.

8.3.2 If the Mahalanobis distance of the sample under test is within the user determined confidence limits, the calibration may be used to estimate the properties in the sample under test. If the Mahalanobis distance of the sample under test exceeds the user determined confidence limits the calibration my not be validly used to estimate the properties in the sample under test.

8.4 The NNMD is used to detect if the measurement of a sample under test falls within a void in the multivariate space occupied by the measurements of the samples in the validation set.

8.4.1 The NNMD is given by:

$$\text{NNMD} = \min\left((\mathbf{x} - \mathbf{x}_i)^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{x} - \mathbf{x}_i)\right) \tag{17}$$

where:

$\mathbf{x}_i$ = the ith column of $\mathbf{X}$, and

min( ) = selection of the minimum value in the resulting vector within the parentheses.

NOTE 8—Here, $\mathbf{x}$ and $\mathbf{X}$ are mean centered by the procedure described in 8.3.1 prior to the calculation of NNMD.

8.4.2 If the NNMD of the sample under test is within the user determined confidence limits, the calibration may be used to estimate the properties in the sample under test. If the Mahalanobis distance of the sample under test exceeds the user determined confidence limits the calibration my not be validly used to estimate the properties in the sample under test.

8.5 The user should determine if the application of the model also requires validation of performance of the measurement qualification techniques described in 8.2-8.4. If such validation is required, the user should determine the types of samples and measurements (anomalous measurements) which should be excluded reliably by the measurement qualification techniques as well as the degree of confidence required for reliable exclusion of such samples. The measurement qualification techniques should be tested on a number of comprehensively representative anomalous measurements sufficient to statistically test that the qualification techniques reliably exclude these anomalous measurements to the required, user specified, degree of confidence.

8.5.1 Anomalous measurements may be caused by transient disruptions to the measurement system, transient problems with sample handling or presentation, the presence of unexpected constituents in a sample being measured, transient disruptions in a process being measured, measurement of an incorrect sample due to clerical error or operator error, samples with constituents or properties outside the range for which the calibration was validated, or any other phenomenon or circumstance which causes the measurement to contain variance which was not adequately represented in the validation samples. This listing of potential causes of anomalous measurements is exemplary, not comprehensive; the user should take into account any conditions or circumstances pertinent to the application.

9. Keywords

9.1 calibrations; empirically derived multivariate calibrations; validation

APPENDIX

(Nonmandatory Information)

X1. EXAMPLE OF PROTOCOLS FOR VALIDATING ANALYTICAL METHODS USED BY THE PHARMACEUTICAL INDUSTRY

INTRODUCTION

This Appendix X1 is for information only. Its inclusion in this practice does not imply any endorsement of the procedures contained herein by ASTM, nor does it suggest that these procedures are applicable outside of pharmaceuticals.

X1.1 The pharmaceutical industry is regulated, both in the United States and in foreign countries, by official governmental agencies. In the United States, the official regulatory agency is the Food and Drug Administration (FDA), whose prime mission is, in the case of pharmaceutical preparations, to ensure the safety and efficacy of those preparations. The FDA has developed guidelines and has recommended guidelines developed by other entities that describe suitable ways to test (evaluate) methods of chemical analysis to ensure the safety and efficacy of the pharmaceutical preparations to which they are applied. Some of the other entities that have developed guidelines that are recommended by the FDA, or whose guidelines have been successfully used in testing and validation of methods of chemical analysis, are the United States Pharmacopeia (USP), the European Agency for the Evaluation of Medicinal Products (EMEA), and the International Conference on Harmonisation (ICH) as well as the American International Society for Testing and Materials (ASTM).

X1.2 To the extent possible, the recommendations in Appendix X1 are intended to be compatible with these already-existing guidelines:

USP Chapter <1119> (**3**)
USP Chapter <1225> (**4**)
EMEA CPMP/QWP/3309/01 (**5**)
ICH Q2 (R1) (**6**)
Practices E1655
Practice E1790

X1.3 Definitions

X1.3.1 *assay*—a measurement procedure that provides a result that is an estimate of the true, although unknown, value of the property measured, and which allows an accurate statement about the sample property being measured.

X1.3.2 *identity test*—test to ensure the identity of the analyte.

X1.3.3 *intermediate precision*—variation within a laboratory, as on different days, or with different analysts or equipment within the same laboratory.

X1.3.4 *library*—a set of various materials of known properties, to which materials of unknown properties may be compared.

X1.3.5 *repeatability*—the precision when independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.

X1.3.6 *specificity*—the ability to assess unequivocally the property in the presence of other properties that may be expected to be present and varying, such as impurities, degradation products, and matrix components.

X1.4 The requirement shared by these documents is to demonstrate that the analytical method under consideration meets a criterion that can be summed up in the short phrase,“ suitable for the intended application.” In other words, application of the analytical method under consideration must guarantee that the pharmaceutical product under test in fact falls within the window of therapeutic activity, delimited at the lower end by the known minimum amount needed for efficacy and at the upper end by the known maximum amount to prevent potential toxicity. It must also guarantee that the pharmaceutical product under test falls within the manufacturer’s specifications for the critical attributes of the product.

X1.5 The data from the proposed analytical method may be subjected to a well-defined mathematical transformation before the validation is performed. If the data is transformed by a transformation of this nature, then the transformation becomes part of the analytical method and should be applied during application of the method.

X1.6 Validation is performed for a combination of a particular, specified instrument; defined measurement conditions; and a particular, specified calibration model.

X1.6.1 Minor changes to a model, for example, a change in the zero adjustment, achieved by changing the constant (b0) term of a model for quantitative calibration may not require a full revalidation. If the applicable SOP permits, a minor change of this sort to the model may require revalidation, for example, of only bias, precision and accuracy.

X1.6.2 A major change to the instrument will generally require a full revalidation. An SOP may be defined which specifies when changes are deemed sufficiently minor that they require lesser revalidation, for example, revalidation only of bias, precision and accuracy.

X1.6.3 It is assumed in Appendix X1 that the instrument used has passed whatever instrument-specific tests are mandated by their IQ, OQ, and PQ requirements (see the recommendations in Ref (**3**) for example requirements for Near Infrared Spectrometers).

X1.6.4 If revalidation for specific cause is not required, quality control (QC) procedures shall be performed at prespecified intervals. As confidence in the accuracy and other critical performance characteristics of the method is attained, the intervals may be lengthened. For example, immediately after initial deployment of the method, QC may be performed daily for a week or two, then weekly for a month or two, then monthly thereafter. QC procedures will generally not require full revalidation; a lesser degree of testing, such as that described in X1.6.1 may be used. The QC procedure, and the intervals at which it is to be applied shall be described and documented.

X1.7 Multivariate methods of analysis are considered alternative analytical methods. Therefore, if a sample fails to meet one or more of the required specifications when tested with the multivariate method, it may be retested using the primary, or accepted reference, analytical method for that material, the result of which will be considered conclusive.

X1.8 Chemists recognize the existence of two generic classes of chemical analysis, qualitative and quantitative analysis. The existence of these two classes of analysis is also recognized by regulatory authorities, who require appropriate validation tests for the two classes. It is therefore necessary to consider the two classes of analytical methods separately and to prescribe validation properties appropriate for each class.

X1.8.1 *Qualitative Analysis:*

X1.8.1.1 In the environment of testing pharmaceutical preparations, qualitative analysis, is also called “identity testing” or “identification”. Conventional identification using chemical or other tests as the reference method is often based on the application of more than one analytical method. Therefore it should be made clear which method (or methods) are replaced by the multivariate method under test.

X1.8.1.2 Qualitative analysis has two main applications, both of which are considered qualitative analysis for the purpose of this appendix.

(1) One application is ascertaining whether a given material is one of a set of known materials, and ascertaining whether a given material is what it is supposed to be.

(2) For the second application the interest is in knowing if the material is what it is supposed to be. If it is not what it is supposed to be, then there may be little or no interest in what it is. This might be the case if the identification test is simply a pass/fail type of test, such as might be used, for example, when testing incoming raw ingredients, or verifying that a final product is within specifications. This is distinct from the applications described in X1.8.1.2 *(1)*, where there is interest in the identity of the material.

X1.8.1.3 A shorthand term for the key critical attribute needed by an analytical method for qualitative analysis is Positive Fraction Identified as described in 3.2.8. To demonstrate specificity the user must demonstrate that the method used has the ability both to recognize when a material under test is of the same type as one of the materials in the training set and identify it as the correct material, and also to recognize when the material under test is not the same as any of the materials in the training set and provide a suitable notification.

(1) To demonstrate that the method can recognize a material under test as being in the training set, the validation samples must include all the variability expected in those samples that are to be recognized as part of the set. An example would be the inclusion of different production batches of the material. These should give a positive identification.

(2) The nature of the samples to be included will depend on the application. For example if material of the same chemical nature but of different particle size, or different polymorphs of the material are to be recognized, then the validation procedure should demonstrate that they are considered of the same type by giving a positive identification.

(3) To demonstrate that the method can recognize that a material under test is not in the training set, the validation process must challenge the method with samples that are different from, but similar to those in the training set. These samples should include materials that are likely to arise during routine application of the model; examples would include reaction by-products, impure samples, homologues of the intended material, etc., also materials received on-site that are similar to recognized materials in visual appearance, chemical structure or name. The method should reject these materials. A case study including a challenge set of samples can be found in Ref (**7**).

(4) Materials of the same chemical type but of different particle size or polymorphic form as recognized materials should also be included among the challenge samples. Unless they are to also be recognized as being of the same type, as described in X1.8.1.3 *(1)*, the method should reject all these materials.

X1.8.1.4 Revalidation of a qualitative method is recommended when the conditions shown in Table X1.1 have changed; these recommendations are in conformance to the requirements of Ref (**5**).

X1.8.1.5 In case of an ambiguous conclusion, the method should be adjusted so that a sample will be unambiguously recognized or rejected. In cases where this cannot be accomplished, the sample must be tested by the original accepted reference analytical method (or methods) used for this identification.

X1.9 Quantitative Analysis

X1.9.1 *Samples*—Section 18 of Practices E1655 contains recommendations for selecting a suitable sample set for validating quantitative calibration models. The validation set may include production samples only, or may include both production and development batches. Validation samples shall not

TABLE X1.1

Change	Revalidation of Library	Revalidation of Instrument Performance	Revalidation of Method
Change to the reference library[A]	YES[B]	NO	n/a
Software changes to instrument[C]	YES[D]	YES	NO
Software changes to analysis computer[C]	YES[D]	NO	YES
Hardware changes to instrument[E]	YES[D]	YES	YES

[A] Addition or deletion of library members, or addition or deletion of samples to a library member.
[B] Depending on the chemometric technique used, revalidation may not be needed when a library member is deleted from the library.
[C] New software or new version of software.
[D] If the change affects computed absorbance values.
[E] For example, upgrade or replacement of lamp, optical or electronic components, instrument, change of instrument location.

have been used in any way during the creation of the calibration model. The validation samples should include variations of the analyte up to the extremes of the range over which the calibration is to be applied. The validation set should also include the variations of other constituents (for example, excipients) as well as in physical properties (for example, particle size, polymorphic form, temperature, etc.) and other critical quality properties expected to be encountered during routine use of the method. The range of variation of all properties should approximate or equal the range of those properties which will be present in the samples on which the calibration is to be applied. As many batches as possible is preferred for validation but in any case the number of batches should not be less than ½ the number of batches used for the calibration set. An attempt should be made to include equal numbers of samples above and below the target value for the analyte.

X1.9.2 *Validation Tests*—The tests needed to validate a method depends on the purposes of the method. Table X1.2 describes the tests needed for models intended for use in various applications.

X1.9.3 *Determination of Analytical Performance Characteristics:*

TABLE X1.2

Analytical Performance Characteristic	Assay of Major Components	Quantitative Minor Component Analysis	Limit Tests for Minor Components	Tests of Performance Characteristics[A]
Accuracy	Yes	Yes	[B]	[B]
Precision	Yes	Yes	No	Yes
Specificity	Yes	Yes	Yes	[B]
Detection limit	No	No	Yes	[B]
Quantification limit	No	Yes	No	[B]
Range	Yes	Yes	No	[B]
Bias	Yes	Yes	Yes	[B]

[A] For example, dissolution profile, drug release rate, etc.
[B] May or may not be required, depending on the nature of the specified test and the application.

X1.9.3.1 *Accuracy*—Accuracy can be determined by comparison of the results from the analytical method under test to those from the accepted reference method, and should be reported as the SEV, calculated according to Eq 5-8. Note that the SEV is sometimes called SEP, although SEV is preferred. The SEV shall not be greater than 1.4 times the Standard Error of the Laboratory unless justified in view of the required accuracy of the test method.

X1.9.3.2 *Precision*—Repeatability should be assessed using a minimum of three replicates of each of a minimum of three different concentrations, one at 100 % of the target value of the analyte, and one at each of the two ends of the range, as specified in X1.9.3.13. The pooled standard deviation of the sets of repeat readings shall be computed and reported. Alternatively, repeatability may be assessed using a minimum of six replicates of a formulation containing 100 % of the target value for the analyte; in this case the standard deviation of the readings shall be computed and reported.

X1.9.3.3 For intermediate precision the effects of random events (for example, different analysts, different days, etc.) on the analytical procedure should be established. The standard deviation and confidence interval should be reported for each type of precision investigated.

X1.9.3.4 The precision should be equivalent to or better than that of the reference method.

X1.9.3.5 *Specificity*—The extent of specificity testing depends on the application of the multivariate analytical method. Lack of specificity can be compensated through the use of other supporting analytical procedures.

X1.9.3.6 Independent samples of substances represented in the training set but not used to create the calibration model (for example, different batches) must be tested and all pass.

X1.9.3.7 Potential challenges should be presented and tested against the model. These challenges should be rejected (that is, no match).

X1.9.3.8 Specificity can be supported through examination of the spectrum of the analyte (either an unmodified absorbance spectrum or modified by one or more well-defined mathematical transformation(s) to enhance spectral features) and comparison of the spectrum with known materials.

X1.9.3.9 Relevant name and substance analogues should be included in the challenge set unless their absence is justified. Examples of valid justifications can be found in Ref **(7)**.

X1.9.3.10 *Detection Limit*—Validation should be verified by preparing or obtaining a suitable number of samples known to be at or near the detection limit. To the extent possible, equal numbers of samples above and below the detection limit should be measured.

X1.9.3.11 Determination of the multivariate detection limit is an area of extensive and ongoing activity. Several approaches to determination of the univariate detection limit based on the final value produced by an instrument and model are presented in Ref **(6)**. We summarize them briefly here:

(1) Based on signal-to-noise ratio: a measured value greater than three times the standard deviation of readings from samples known to not contain analyte (that is, a blank) indicates the presence of analyte.

(2) Based on response and slope: a graph of test results versus true (that is, accepted reference) values may be drawn and the formula applied: $DL = 3.3s/S$, where s is the SEV of the graph and S is the slope of the relationship between test results and true (that is, reference) values. The standard deviation of the Y-intercept may also be used for s.

(3) The detection limit and the method of determining it should be documented.

(4) Limit tests substantiate the fact that the amount of analyte is above or below a certain level. The detection limit is usually expressed as the concentration of analyte (for example, percentage, parts per billion) in the sample.

X1.9.3.12 *Quantification Limit*—Validation should be verified as for detection limit, by preparing or obtaining a suitable number of samples known to be at or near the quantification limit. To the extent possible, equal numbers of samples above and below the quantification limit should be measured. Recommendations for determining the detection limit are in Refs (**4, 6**). A graph of test results versus true (that is, reference) values may be drawn and a formula similar to the one for detection limit may be applied: $DL = 10s/S$, where s is the SEC of the graph and S is the slope of the relationship between test results and true (that is, reference) values. The standard deviation of the Y-intercept may also be used for s. The quantification limit and the method of determining it should be documented.

X1.9.3.13 *Range*—The necessary range for validation shall depend on the purpose of the analysis. If for assay, the range shall extend from 80 % to 120 % of the target value of the analyte. If for content uniformity, the range shall extend from 70 % to 130 % of the target value of the analyte. If 120 % or 130 % of the target value for the analyte requires samples containing more than 100 % absolute of the analyte, then the required upper limit of the range shall be 100 % absolute. If samples containing 100 % absolute amount of analyte cannot be made (for example, tablets spontaneously disintegrate) then the upper limit of the range shall be set equal to the maximum amount of analyte that allows a suitable sample preparation to be made.

X1.9.3.14 *Bias*—The bias should not differ from zero by more than three times the SEV divided by the square root of the number of independent samples in the validation set.

X1.9.3.15 Effects of possible, relevant variations, such as temperature (environmental and sample), humidity, position of sample in the instrument, sample container, probe depth and applied pressure, should be understood and documented. Instrumental variation (for example, replacement of lamp, optical reference standard, electronic or optical components, etc.) can also be considered in the validation. This listing of potential items for consideration is exemplary, not comprehensive; the user should take into account any external conditions pertinent to the application.

REFERENCES

(1) Kramer, R., *Chemometric Techniques for Quantitative Analysis*, Marcel-Dekker, New York, NY, 1998.

(2) Martens, H., and Naes, T., *Multivariate Calibration*, John Wiley & Sons Inc., Hoboken, NJ, 1989.

(3) U.S. Pharmacopeia and National Formulary <1119>, "Near Infrared Spectrophotometry," USP 29–NF 24, 2006.

(4) U.S. Pharmacopoeia <1225>, *Validation of Compendial Methods*, Vol 24, 1999, pp. 2149–2152.

(5) The European Agency for the Evaluation of Medicinal Products, CPMP/CVMP, "Note for the Guidance on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations," Feb. 20, 2003.

(6) ICH Q2 (R1), *Validation of Analytical Procedures: Text and Methodology,* November 2005.

(7) Plugge, W., and van der Vlies, C., *Journal of Pharmaceutical and Biomedical Analysis*, Vol 11, No. 6, 1993, pp. 435–442.